The Manager's Guide to Statistics

Erol A. Peköz

2020 Edition

Printed in the United States of America.

ISBN 978-0-9795704-4-5

To order give your college bookstore the information below:

The Manager's Guide to Statistics, by Erol A. Peköz (2020), ISBN 978-0-9795704-4-5, email: pekoz@bu.edu.

Download data sets at the following address:

`http://people.bu.edu/pekoz/bookdata.zip`

Contents

Preface

This book is an introduction to statistics for someone who does not need to know all the details of statistical theory and would just like to know how statistics is commonly used in business. We cover some of the basic statistical tools, some pitfalls to watch out for when using them, and we give some intuitive explanations of how everything works. We try to explain everything in words rather than with Greek symbols or mathematical formulas. You'll also find the technical details and notation in technical notes so as a student you could be comfortable using other statistics books in the future.

It's difficult for students of statistics to learn, at the same time, the imposing mathematical notation and the subtle, fragile concepts behind the scenes. If a teacher tries to teach both at the same time, invariably attention paid by students to the concepts goes out the window. The student leaves the course with only a vague recollection of a jumble of Greek letters and of some frustrating time spent on the computer pointing and clicking and coping with error messages. This is why we make such an effort to teach the concepts first without the imposing notation and then include much of the notation and technical details as technical notes afterwards.

We recommend using Excel or R for the statistical calculations. Excel and R are ubiquitous in business and becoming more comfortable with them is of great value in itself. Though Excel is not the program of choice for professional statisticians, it is a program of choice for business people.

This book is designed to be the textbook for a one semester introductory course for undergraduate business students, MBA students, or other aspiring leaders and decision makers.

About the Author

Erol A. Peköz is a Professor in the Questrom School of Business at Boston University. He received his B.S. from Cornell University and Ph.D. in Operations Research from the University of California, Berkeley and has published a number of technical articles in applied probability and statistics. He also has taught at the University of California, Berkeley, the University of California, Los Angeles, and Harvard University. At Boston University he was awarded the 2001 Broderick Prize for Teaching, and is also a co-author of the textbook *A Second Course in Probability* (with Sheldon Ross).

Acknowledgements

I would like to thank Paul Berger, Andrew Gelman, Janelle Heineke, Barry Kadets, Mark Kean, Sheldon Ross, Michael Shwartz and Mustafa Yilmaz for their valuable contributions.

Chapter 1

Seeing the Real Story

Statistics are tools for seeing and telling a story from data. Computers can usually handle the number-crunching calculations we may need, but interpreting the output and seeing a story is up to us. Instead of starting off by teaching you how to do calculations, we will start with some advice on how to see the real story behind the numbers.

One theme throughout this book is that it is easy to be misled by statistics if you don't know what to watch out for. Even honestly gathered data may appear to be telling one story on the surface, when actually quite the opposite is the real story. But after this chapter you shouldn't just become skeptical of all of statistics. Being overly skeptical is just as bad as being overly gullible: both keep you from the real truth. Our goal for you is to become wise enough to know when to be skeptical and when to believe. In this chapter we will cover a few commonly encountered ways statistics tend to mislead people, so you will know what to watch out for.

1. Comparing Rates: Missing Denominators

Sometimes the ratio of two numbers tells more than either of the numbers do by themselves. Next are a few examples.

How many hospital beds does a city need? If your city has ten times more hospital beds than another city, does this mean your city probably has

more beds than it needs? To start to answer this, we really need to know the number of people in each city. We should not just compare the **count** of beds, but should instead compare the ratio of the number of beds to the number of people in the city. We call this type of ratio a **rate**. For example, if a city of 200,000 people has 100 hospital beds, this corresponds to a rate of one hospital bed for every 2,000 people. Though assessing if a city has more beds than it needs is controversial and complex,[1] we should definitely start off by looking at rates instead of counts.

> Watch out if someone is comparing counts when they should instead be comparing rates.

Are you safer without a seat belt? During the year 2002 in California, 1,524 people wearing seat belts were killed in car accidents but only 1,343 people without seat belts were killed in car accidents.[2] Do these numbers mean you are safer without a seat belt?

Solution. No, you are not safer without a seat belt. Since most people wear seat belts, we would expect a large number of deaths among people wearing seat belts. To see the benefits of a seat belt you should compare the death rates for the people with and without seat belts. You could do this by dividing the number of deaths for each group by the total number of people (or the total number of accidents) for each group. This rate would turn out to be much higher for the group of people who weren't wearing seat belts. It's also true that many more people are killed riding in cars than riding motorcycles: only 318 motorcyclists were killed in 2002 in California. Even though motorcycles are more dangerous than cars (and the death rate turns out to be higher), there are fewer deaths because there are fewer motorcyclists.

Is New York City more dangerous than Iraq? The death rate for Americans (that is, deaths per thousand Americans per year) in Iraq during the

[1] There is research that claims that the over-supply of hospital beds induces demand for them. A study by E. Peköz and M. Shwartz funded by the *Department of Health and Human Services, Agency for Healthcare Research and Quality* titled "Do More Hospital Beds in an Area Induce Excess Demand?" is investigating the evidence for this.

[2] See *2002 Annual Report of Fatal and Injury Motor Vehicle Traffic Collisions* at `http://www.chp.ca.gov/pdf/2002-sec4.pdf`, page 21.

war was actually lower than the death rate in New York City during the same time period.[3] Does this mean it was safer to be sent to Iraq than to New York City?

Solution. No, Iraq was much more dangerous. It's true there were fewer American deaths per year in Iraq than in New York City during that time, and this is because there were more Americans in New York City than in Iraq. But that still doesn't explain why there were fewer deaths per thousand Americans in Iraq. The reason for this difference in rates is because Americans in Iraq had a different age range than people in New York City. In Iraq the Americans were primarily young healthy soldiers, while the New York City death rate included the sick and the elderly, groups that typically have high death rates. If you compared Americans in Iraq with people of the same age range in New York City you would find a much higher death rate in Iraq. This type of age difference is called a *confounding factor*; we will talk more about these types of factors in the next section.

Exercises

1. **Workplace safety.** Managers of a manufacturing plant keep track of on-the-job accidents. Last year there were 50 worker injuries that happened on the day shift and only five worker injuries that happened on the night shift. Does this mean the night shift was safer? Or are we comparing the wrong type of numbers?

2. **Automobile theft.** Insurance industry records show the Cadillac Escalade SUV has the highest theft rate of all cars in terms of the number of thefts per thousand vehicles. But in the same records if you just look at the total number of thefts, more Toyota Camrys are stolen each year than any other car—including Escalades.[4] The 1989 Camry, in particular, is the one most often stolen.

[3] The New York City mortality rate was around 700 per 100,000 people per year as measured in the year 2000 census. The article "Counting the Dead" by James Dunnigan posted on July 29, 2004 on strategypage.com reports the figure as 360 per 100,000 troops per year in Iraq. The yearly death rate in New York City for men 20-24 was 120 per 100,000 in 1999-2001 (it was only 40 per 100,000 for women, and for men over 85 it was 15,000 per 100,000). See `http://strategypage.com/dls/articles/200472922.asp`, and `www.nyc.gov/html/doh/pdf/vs/2002sum.pdf`.

[4] `http://www.auto-theft.info/Statistics.htm` and `http://money.cnn.com/2004/02/27/pf/autos/nicb_most_stolen/`

(a) How can you explain the difference here?
(b) If you are considering buying either a Camry or an Escalade, which car would be more likely to get stolen?

Which would a thief prefer?[5]

3. **Pharmacy errors.** A newspaper article reports about a dangerous surge in pharmacy prescription errors.[6] The article details how the vast majority of complaints to the Massachusetts Department of Public Health have been lodged against CVS Corporation's pharmacies. The article goes on to say this is particularly troublesome because CVS is one of the larger pharmacy chains in the state. Does this mean CVS pharmacies are having problems? Or are there different numbers we should be looking at?

2. Comparing Two Groups: Confounding Factors

It is typical for researchers to compare two groups of subjects they are studying and try to draw conclusions based on differences they see. To study the effectiveness of a drug, for example, medical researchers may compare outcomes for people who take the drug with people who don't take the drug. To test the effectiveness of an advertising campaign, market researchers may compare sales for two regions where two different advertising campaigns are used.

This may seem like a straightforward process, but it can be complicated by the presence of **confounding factors** (also called **confounders**). A confounding factor is an important difference between the two groups you are

[5] Photos from `http://www.geartekcorporation.com/dailyphoto/2005/toyotacamry.html`, and `http://www.cadillacforums.com/cadillac-models/cadillac-escalade.html`

[6] "Massachusetts pharmacist woes a prescription for peril," by Jessica Heslam, *The Boston Herald,* Thursday, July 14, 2005, page 2.

comparing, other than the one you're primarily interested in. Such a difference is important if it has a big impact on what you are measuring in the study.

In the example of the previous section where we compared Americans in New York City with Americans in Iraq, we were primarily interested in the locations of these two groups of people. The confounding factor—the other important difference—was the age difference between the two groups of people. Overlooking this confounding factor at first is what gave us the misleading conclusion that Iraq was safer than New York City. Of course the two groups also differed in that most Americans in Iraq were US government employees, but this difference is not as important as the age difference and probably would not be a confounding factor.

> A confounding factor is a characteristic that differs between two groups you are comparing, other than the one you're primarily interested in. This characteristic also should have a big impact on what you are measuring in the study.

How can you avoid confounding factors? One way is to recognize them and then try to arrange the groups you compare so you don't have confounding factors. For example suppose, as in the example of the previous section, the ages of people are very different in two groups you would like to compare and you believe that age will have an important impact on what you are primarily interested in. You can then divide the groups up into several smaller groups and then only compare groups of people with the same age. This process in this case is called **adjusting for** age or **controlling for** age.

Sometimes it can be difficult to adjust for confounding factors. For example, consider a study of the harmful health effects of smoking. Even though people who smoke tend to be less healthy than people who don't smoke, people who smoke as a whole tend to drink more than people who don't smoke. It's difficult to tell if it's smoking causing health problems, or if it's drinking. This type of study where researchers compare two naturally formed groups of people is called an **observational study**.

Another way to avoid confounding factors is to conduct a **randomized experimental study**. In a randomized experiment, researchers randomly divide

subjects into a **treatment group** and a **control group** (a group that receives no treatment or an inactive **placebo** treatment such as a sugar pill). Then the researchers apply some form of treatment to the first group and compare the results of the two groups. Although there will always be some differences from person to person in any study, randomly dividing the subjects into two large groups usually fairly balances out any potential confounding factors so they don't cause problems.

Usually experimental studies are preferable to observational studies, but important decisions must very often be made on the basis of observational studies. Experimental studies can be too expensive or impractical to conduct. Human studies on smoking, for example, are inherently observational studies. In our country it is not possible to force people to smoke for twenty years just to compare their health to people who didn't smoke.

Is your company selling a dangerous product? Shortly after a popular blood pressure drug entered the marketplace, a study of hospital records showed that people who had taken this drug subsequently had a higher rate of heart attacks than people who took other blood pressure drugs.[7] Did this mean the new drug was dangerous? Or was there a likely confounding factor to consider?

Discussion. There was a confounding factor. This was a high blood pressure drug intended for the more serious blood-pressure cases so people taking this drug tended to have more heart problems even before they started taking the medication. The prior health history of the subjects in this study would therefore be an important confounding factor, since it was very different between the two groups. Randomized experiments showed that the drug in fact was safe.

Is the market for a drug secure? During the 1990s Wyeth Pharmaceuticals was selling millions of hormone treatments to women annually. Large

[7] B. M. Psaty, S. R. Heckbert, T. D. Koepsell, D. S. Siscovick, T. E. Raghunathan, N. S. Weiss, F. R. Rosendaal, R. N. Lemaitre, N. L. Smith et. al., "The risk of myocardial infarction associated with antihypertensive drug therapies," *Journal of the American Medical Association*, Vol. 274 No. 8, August 23, 1995, and Robert M. Califf and Judith M. Kramer, "What Have We Learned From the Calcium Channel Blocker Controversy?" *Circulation*, 1998;97:1529–1531, and `http://cumc.columbia.edu/news/journal/journal-o/archives/jour_v17n1_0009.html`

observational studies over many years unquestionably showed that women who took these treatments after menopause had lower rates of heart disease than women of the same age who didn't take the treatments.[8] Pharmaceutical industry financial analysts believed the market for these hormones was secure, and the stock price of Wyeth was soaring. Then in 2002 a small randomized experimental study came out that surprisingly found that hormone therapy increased the risk of heart disease—the very problem it was supposed to prevent. Imagine you are an investor in Wyeth pharmaceuticals just learning this news. Should you trust the earlier large studies or should you rush to sell your stock?

Discussion. You should have rushed to sell your stock because the earlier studies had reached the wrong conclusion. Those studies were observational studies, which means that researchers just looked at the health records of people who had been prescribed hormone treatments by their doctors. These women weren't randomly selected for the treatment as they would be in an experimental study. Researchers now believe the type of women who sought out and stuck with hormone replacement therapy were the type of women who tended to get better health care overall, tended to lead more health-conscious lives and tended to be healthier to begin with than women who weren't taking the therapy or who didn't stick with it. Also, it is believed doctors were less likely to prescribe this therapy for women with risk factors for heart disease. The prior health of these women was a confounding factor, since it was very different between the two groups being compared. Although researchers at the time had tried, they weren't able to properly adjust for all these factors in the observational studies.

In the experimental study, researchers assigned volunteers randomly to different groups and so they avoided the confounding factors that existed in the observational studies. Researchers now see how the observational studies were very misleading and hormone replacement therapy is now not recommended for prevention of heart disease—though there are recent reports of benefits observed in some age groups that weren't well documented before.[9]

[8] "The WHI Estrogen Alone Trial—Do Things Look Any Better?," by S. Hully, and D. Grady, *Journal of the American Medical Association*, April 14th, 2004, volume 291, No. 14, Pages 1769–1771. See also Hersh AL, Stefanick ML, Stafford RS, "National use of postmenopausal hormone therapy: annual trends and response to recent evidence," *Journal of the American Medical Association*, 2004; 291: 47–53.

[9] "The Latest Wisdom on Hormone Therapy," by Katherine Hobson, 4/20/07, *US News and World Report*, `http://www.usnews.com/usnews/health/articles/070420/`

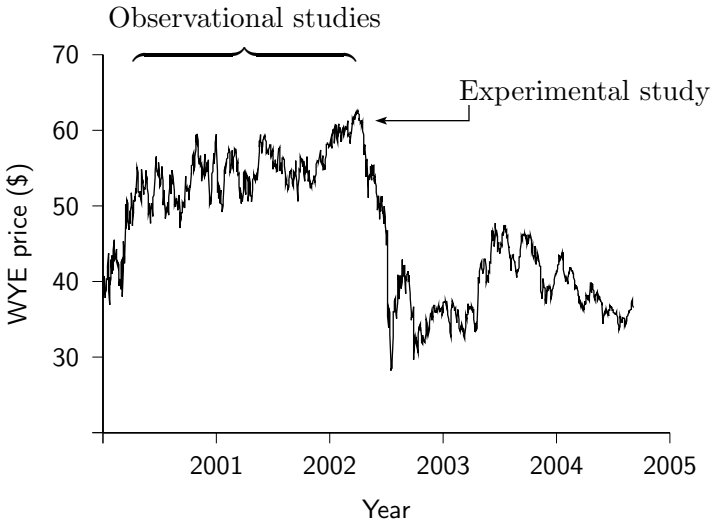

Figure 1.1. Wyeth pharmaceutical company's stock price plummets after an experimental study finds that hormone therapy – one of its most important products – increases the risk of heart disease.

This is an example of how a well-designed experimental study can overturn a conventional belief.

Incidentally, Wyeth's stock price (ticker symbol WYE) took a sharp nose dive shortly after the news of the experimental study came out (see Figure 1.1). This shows that as an investor or industry analyst, it pays to understand the difference between experimental studies and observational studies and the problems confounding factors can cause.

Can you have confounding factors in a randomized study? A study that looks at first as though it is experimental may actually be observational, and may be therefore susceptible to confounding factors. If any of the subjects in a study influence whether or not they are in the treatment group or the control group, the study is no longer experimental: it becomes an observational study. For example, researchers testing the effectiveness of a new drug with side effects may divide volunteers randomly into a treatment

`20health.hormone.htm` and "How NIH Misread Hormone Study in 2002," By Tara Parker-Pope, *The Wall Street Journal*, July 9, 2007, Page B1, `http://online.wsj.com/article/SB118394176612760522.html`

group and a control group. So far it looks as though the study is a randomized experiment, but almost always some of the volunteers drop out of the study in the middle. It may be tempting at the end of the study to ignore these dropouts, but this would introduce a confounding factor. If the drug has side effects, people who drop out are likely to be the ones who can't tolerate the side effects and so people who remain in the study are likely to be the strongest and healthiest members of the treatment group. If we allow subjects themselves to decide whether or not they drop out of the treatment group, the study becomes observational rather than experimental. In order to avoid this problem, the researchers should still consider people who drop out of the treatment group as part of the treatment group. The entire treatment group, including dropouts, should be compared with the entire control group, including dropouts. This way the study remains experimental and we will not have confounding factors.

When we detect an association between a treatment and a result it is tempting to conclude that the treatment causes the result. But in an observational study just because two things are associated with each other doesn't mean one causes the other. In fact, the causation may even go in the other direction. We will illustrate this next.

Association is not always the same as causation.

Is it dangerous to quit a hazardous job? Researchers studied high-voltage electrical utility workers and divided them into two groups: people currently on the job and people who had recently quit working.[10] It turned out that the first group was healthier than the second group. Should we conclude from this that, even though working near high-voltage wires may be hazardous, quitting such a job is even more hazardous?

Discussion. We should first ask if the age or gender mix in the two groups was significantly different. Most likely the former workers were older than current workers, and this could be a confounding factor. To handle this difference, we can adjust for age by making a comparison for each age group separately. This means we compare young people in both groups, then com-

[10]Richardson et al., "Time Related Aspects of the Healthy Worker Survivor Effect," *Annals of Epidemiology,* 14:633-639, 2004.

pare middle-aged people in both groups, and so forth. Interestingly, the former workers were still less healthy even after adjusting for age.

In this study it turned out that many people quit their jobs because they were having health problems. This means poor health caused them to quit, and it was not quitting that caused the poor health. When it appears one thing may be causing another thing to happen, the causation may actually be running in reverse.

Incidentally, researchers also looked at the group of people who had quit their job more than five years ago. Surprisingly this group was healthier than both the other groups. Does this mean that high-voltage exposure could be good for you in the long run? No, unfortunately the sickest people who quit their job because of poor health did not survive five years. Those who survived had to have been extraordinarily healthy people to begin with. This was a way to effectively select out only the healthiest people, and is sometimes called **selection bias** or **survivor bias**. The next section gives more examples of this phenomenon.

Exercises

4. **Prison education programs.** Massachusetts is considering a new statewide education program to better rehabilitate prisoners before their release with the hope of reducing the percentage who will be back in prison. This type of program has already been tried on a voluntary basis in one county where courses in anger management, addiction, and post-release planning were made available to inmates six months before their release. A newspaper article titled "Prison education programs cut rate of reoffending" details how inmates who participated in these courses were half as likely to re-offend as the ones who didn't participate.[11]
   (a) Is this study observational or experimental?
   (b) Do these results mean the counseling program is effective or is there an important confounding factor to consider here? Explain.
   (c) How could you conduct a better study where participation is still voluntary?

5. **Car alarms.** A newspaper article titled "Ban car alarms? They disrupt life & don't work" argues that, contrary to popular belief, car alarms

[11] "Prison education programs cut rate of reoffending," *The Boston Globe*, Tuesday, July 13, 2004, page A13, by Joseph McDonough.

do not effectively deter theft.[12] The article refers to a study that finds that car models that come with alarms installed are actually more likely to be stolen than models that don't come with alarms. Does this mean car alarms don't work or is there an important confounding factor to consider here?

6. **Corporate funded studies.** Many drug safety research studies are sponsored by pharmaceutical companies that would financially benefit if the results of the study are favorable.[13] Is this an example of a potential confounding factor?

7. **Antidepressants and suicide.** A high profile study finds that the use of antidepressants significantly increases the risk of suicidal behavior in adults under 25 while it decreases the risk of such behavior in older adults.[14] The study involved over 100,000 people and found that adults under 25 taking antidepressants were more than twice as likely as those taking a placebo to attempt, prepare for, or commit suicide. Someone criticizes the study saying that people who take antidepressants are generally more depressed to begin with compared with people who don't—and this could be a serious confounding factor. Is this valid criticism of the study?

8. **Study volunteers.** Since patients who volunteer to participate in a randomized experimental drug study tend to have health problems to begin with, they don't usually represent the general population. Is this an example of a confounding factor? Explain why or why not.

[12] "Ban car alarms? They disrupt life & don't work," by Aaron Naparstek, *New York Daily News*, Tuesday, July 8th, 2003.

[13] See "Financial ties to industry cloud depression study," Tuesday, July 11, 2006, by David Armstrong, *The Wall Street Journal* `http://www.post-gazette.com/pg/06192/705022-114.stm` and "JAMA says docs misled over industry ties," July 18, 2006, by Lindsey Tanner, The Associated Press. See `http://www.abcnews4.com/news/stories/0706/345549.html`

[14] "Study Finds Medication Raises Suicide Risks in Young Adults," by Benedict Carey, December 6, 2006, *The New York Times*, `http://www.nytimes.com/2006/12/06/health/06drug.html?ex=1323061200&en=d727e889c3f935bc&ei=5090partner=rssuserland&emc=rss`

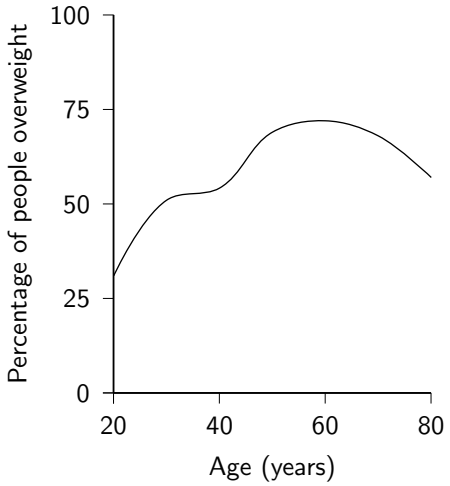

Figure 1.2. The percentage of people overweight versus age. Notice the decrease in the graph after age 60. Does this mean people are shaping up?

3. Selection Bias and Survivor Bias

Did you know that former astronauts have significantly fewer health problems than other people their age? Does this mean space travel is good for your health? Of course not, because one of the criteria for astronauts is that they should be in excellent health to begin with. This is an example of **selection bias**. Next are a few more examples.

Do people shape up as they age? The percentage of people who were overweight was computed for different age groups in a large representative sample of people interviewed during a study in 1997.[15] As a graph of the results in Figure 1.2 shows, the percentage appears to rise until around age 60 and then starts to fall. Does this mean people tend to get in better shape after age 60?

Discussion. People who are overweight generally do not live as long as people who are not overweight. This means the older people are less likely

[15]Data from Department of Health, Social Services and Public Safety, *1997 Northern Ireland Health and Social Wellbeing Survey.* See Table 5.5 in `http://www.dhsspsni.gov.uk/publications/archived/2001/revised_health\%20_lifestyle_report.pdf`

to be overweight. It may appear that people are getting in shape, but it's actually that the overweight people are dropping out of the sample. This form of selection bias is sometimes called **survivor bias**.

There is an important difference in the conclusions you can draw from **longitudinal** data, where subjects are followed over a long period of time, and **cross-sectional** data, where data are gathered at a single point in time. This study was cross-sectional. If the same subjects were instead followed over time, we would see that people do not tend to shape up after age 60. Cross-sectional data are usually much cheaper and faster to gather than longitudinal data, but one has to be cautious about drawing longitudinal-type conclusions from cross-sectional data. In some cases it can be reasonable to make longitudinal conclusions from cross-sectional data, but you should be very cautious when doing this.

> If a study claims things will happen over time, ask if the data are cross-sectional or longitudinal.

Mutual fund historical performance. A mutual fund company shows the historical performance of all the funds they currently sell. Over the past five years every single one of these funds performed above the Standard & Poor's 500 index (S&P 500 index), a commonly used benchmark for investment performance. Does this mean that someone who bought a fund from this company five years ago would probably have beaten the S&P 500 index?

Discussion. Mutual fund companies commonly eliminate, merge, or rename funds that don't perform well. The track records of these funds are erased from the databases that companies and the media generally use to measure performance. Funds that are currently advertised for sale (the "live" funds) are the ones that happened to have performed well. The unlucky people who invested in funds that were eliminated, merged, or renamed (the "dead" funds) very likely did not earn returns above the S&P 500. Eliminating these "dead" funds from the records creates a **survivor bias** and what people think of as the average mutual fund's return is greatly overstated. A recent *Wall Street Journal* article discusses how this bias can be upwards of several percentage points.[16]

[16] "Mutual Fund Math Puts a Sheen on Returns," by Ian McDonald, *The Wall Street Journal*, Friday Jul. 23rd, 2004, p. C1.

This survivor bias not only applies to people and mutual funds, but to publicly traded companies as well. When a company's stock performs very poorly, the company may itself go out of business. This means that the historical performance of currently traded stocks tends to look a lot better than for the average stock. Such survivor-biased historical data can also cause other more subtle problems when you use it in more complicated analyses.

> Survivor bias can appear in surprising places.

How difficult is it to find a job? A university interested in knowing how difficult it is for recent graduates to find a job sends out a survey form to a random selection of graduates a few months after graduation. When they don't get many responses, they decide to keep sending out survey forms to more people until they get a reasonable number of responses. Is this a sensible approach for increasing the number of responses?

Solution. No. Survey researchers have long realized that the type of people who take the time to fill out survey forms usually differ significantly from the type of people who don't, and only using their responses biases the survey. People who are poorly motivated will have difficulty finding a job and will also be less inclined to respond to a survey and, on the other hand, people who have recently started a new job may not have the time to fill out a survey form.[17] These are examples of what is called **non-response bias**. One way to reduce this bias is to contact the people who didn't initially respond either by telephone or repeated mailings or use other ways to induce them to respond. Sending out survey forms to new people simply recreates the same bias on a larger scale.

If someone tells you about the results of a survey they have conducted, don't just ask them how many responses they received. Also ask how many survey forms were sent out.

> A survey with a huge number of responses can still suffer from non-response bias.

[17] See Gerard J. van den Berg, Maarten Lindeboom & Peter J. Dolton, 2004, "Survey Non-response and Unemployment Duration," Tinbergen Institute Discussion Papers 04-094/3, Tinbergen Institute. `http://www.tinbergen.nl/discussionpapers/04094.pdf`

Initial public offering (IPO) brokerage firms. A study of a developing country's stock exchange showed that companies employing prestigious American investment banking firms to handle their IPOs tended to be better respected by investors after the IPO and had stock prices that rose to higher levels. The researchers concluded from this that the prestige of these investment banks can create value even beyond the cost of their fee. Can you think of another explanation?

Discussion. Their conclusion may be true, but this study doesn't effectively justify it. Companies that have money to hire a prestigious firm are certainly more likely to be in better financial shape than companies that don't have the money. Also, prestigious brokers are more likely to be willing to work with companies that are in better financial shape. Be cautious in concluding that prestigious investment banks can add value beyond the cost of their fee based on this study. An important confounding factor here is the financial strength of the company before the IPO; a strong company is more likely to be able to get a prestigious broker and is also more likely to be respected by investors on its own. It's also true that companies with expensive office furniture tend to be more successful than other companies, but it doesn't mean that the furniture itself is responsible. The successful companies are more likely to be able to afford such furniture.

Exercises

9. **Online banking.** A bank considering eliminating tellers and switching completely over to on-line banking services would like to know if its customers would welcome the idea. The bank conducts a survey on its web site and finds that almost everyone says they would welcome the idea. Does this mean the switch would be welcome or is there an important source of bias in this survey? Explain.

10. **Team-building and profitability.** A study shows that companies that send employees to expensive weekend team-building workshops in exotic locations tend to be the most profitable companies in their industries. Does this mean these programs work to help create a profitable company? Or is there a likely confounding factor to consider here? Explain.

11. **Hospital quality.** Massachusetts is considering a controversial proposal to publicize heart surgery survival rates for individual hospitals and

physicians so patients can be better informed consumers.[18] Insurance analysts profiling the performance of two hospitals find that the first hospital has a much higher survival rate for heart surgery than the second hospital, even though the second hospital has better equipment and more experienced and specialized surgeons. Does this mean the first hospital is doing a better job? Or is there a likely confounding factor to consider here? Explain.

4. Descriptive and Inferential Statistics

In this chapter we frequently made statements such as "a study shows that..." or "some data show that...," but we haven't really discussed yet how statistics can be used to draw such conclusions from data. The field of statistics can be divided into two broad categories: **descriptive** statistics and **inferential** statistics. Descriptive statistics covers methods for describing data you have gathered, and inferential statistics covers methods for using the data you have gathered to draw a conclusion about something unknown. For example, suppose we walk into a store and see only women customers. If we just reported we saw only women it could be classified as part of descriptive statistics. If we then used this to say we believe this store generally tends to have primarily women customers, it could be classified as part of inferential statistics. In this book we first study descriptive statistics and then we move on to inferential statistics.

5. Mathematical Summary

Mathematicians use special notation to explain statistical ideas and it's not always necessary to know the notation to understand the concepts. Sometimes, though, it can come in handy to know the notation when communicating technical details with other technical people. Here we summarize the concepts of this chapter using statistical notation.

Given a group of subjects and a factor A, let $X(i, A) = 1$ if subject number i has factor A and let $X(i, A) = 0$ otherwise. For example, three such factors

[18] "Heart surgery data may go public: State looking at patient death rates for doctors," By Liz Kowalczyk, *The Boston Globe*, July 4, 2006. `http://www.boston.com/yourlife/health/diseases/articles/2006/07/04/heart_surgery_data_may_go_public/`

for people could be smoking, drinking, and good health. Given factors A, B, and C let

$$P(A|B) = \frac{\sum_i X(i,A)X(i,B)}{\sum_i X(i,B)}$$

denote the fraction of subjects with factor B who also have factor A (read as "the probability of A given B"), and also let

$$P(A|BC) = \frac{\sum_i X(i,A)X(i,B)X(i,C)}{\sum_i X(i,B)X(i,C)}$$

denote the fraction of subjects with factor B and C who also have factor A (read as "the probability of A given B and C").

We say that C is a **confounding factor** for A and B if both $P(A|B) \neq P(A|BC)$ and $P(B|A) \neq P(B|AC)$ hold. This means the relation between A and B is different whether or not you take factor C into account. If C is a confounding factor, it's possible to see a situation where $P(A|B) < P(A) < P(A|BC)$, meaning the direction of the impact of B on A can be reversed by C.

6. Summary

Subtle sources of bias can leave an impressive-looking statistical analysis with major flaws. Being better able to recognize these situations will help you defend yourself from being misled. Some key concepts to have in mind are the following: comparing counts versus rates; distinguishing experimental and observational studies; identifying confounding factors; adjusting or controlling for a factor; recognizing association and causation, survivor bias, selection bias, and non-response bias.

7. Exercises

12. **On-the-job fatalities.** Government statistics for on-the-job fatalities in 2003 show there were 1,033 on-the-job deaths of construction workers as well as 1,388 transportation workers, 312 police and fire-fighters, 282 factory workers, and 630 management and executive-level workers.[19] Does

[19] `http://www.bls.gov/news.release/cfoi.t03.htm`

this mean management and executive jobs may be safer than construction jobs, but may actually be more dangerous than fighting fires or working in a factory? Or are we comparing the wrong numbers here?

13. Dangerous occupations. Below are some data from 2005 for on-the-job deaths in some dangerous jobs.[20] Which job seems the most dangerous? Which seems the least dangerous? Explain.

Occupation	Total deaths	Total employed
Driver/sales workers and truck drivers	993	3,412,370
Farmers and ranchers	341	829,680
Construction laborers	339	1,493,390
Miscellaneous agricultural workers	176	758,620
Aircraft pilots	81	121,070
Logging workers	80	86,110
Fishers and fishing workers	48	40,540
Electrical power line installers/repairers	36	110,090
Structural iron and steel workers	35	62,940
Refuse and recyclable material collectors	32	73,050

14. Travel-related illness. A newspaper article reports on a study of international travel-related illnesses, a topic of major interest in the tourism industry.[21] A diagram in the article, shown in Figure 1.3, gives the number of people who returned home with health problems after traveling to various regions of the developing world. Of the 15,704 cases of travel-related illnesses reported during the period of the study, 1,115 were from travel to the Caribbean, 1,326 from Central America, 1,675 from South America, 4,524 from sub-Saharan Africa, 2,403 from South Central Asia, 2,793 from Southeast Asia, and 3,517 from other regions.

(a) What story do the data tell about where the most hazardous regions are? Explain.

(b) A few days later a reader wrote in to criticize the diagram and the newspaper subsequently published a correction note. Can you guess what the reader's complaint was?

[20] "America's most dangerous jobs," by Les Christie, *CNNMoney.com*, August 16 2006: 6:19 PM EDT, `http://money.cnn.com/2006/08/16/pf/2005_most_dangerous_jobs/index.htm?cnn=yes`

[21] "The souvenir no tourist wants: dengue fever," February 6, 2006, Tina Cassidy, *The Boston Globe*, and also see the article "Spectrum of Disease and Relation to Place of Exposure among Ill Returned Travelers," Freedman D. O et al., *New England Journal of Medicine,* 2006; 354:119-130, Jan 12, 2006.

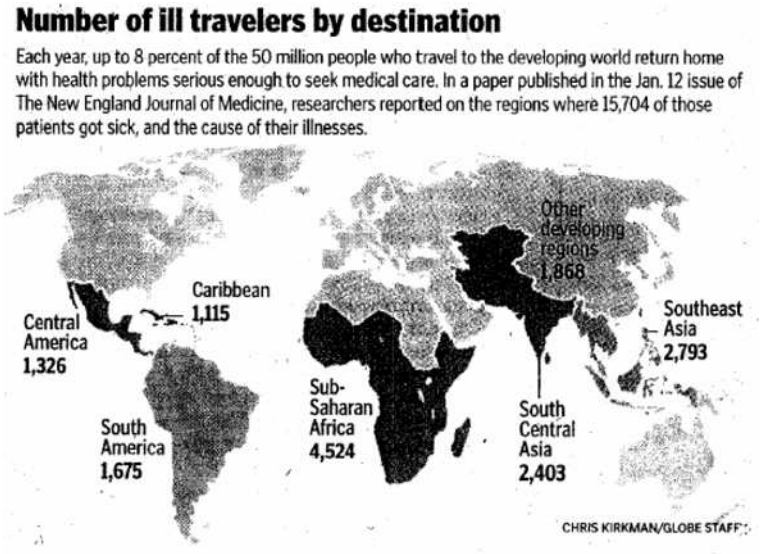

Figure 1.3. The number of people returning home with health problems after traveling to various regions of the developing world. What story does this graph tell about where the most hazardous regions are?

15. Decision making. A recent *Harvard Business Review* article titled "When to Sleep on It" discusses the usefulness of deliberating a business decision.[22] The article describes a study in which subjects were asked to make a number of decisions and each subject was given the option to decide immediately or deliberate while performing an unrelated task. The researchers found that subjects who answered immediately made the best decisions and that "the longer our participants thought about their answers, the more likely they were to include irrelevant information at the expense of relevant information." The article concludes that "conscious deliberation, however long and careful, can be a surprisingly crude and ineffective tool."

(a) Is this study observational or experimental?

(b) Can you think of a confounding factor here?

(c) If you could re-design this study, how would you do it?

[22] "The HBR Breakthrough Ideas for 2007: When to Sleep on It," by Ap Dijksterhuis, February 2007, *Harvard Business Review* pp. 2–54.

16. Online course materials. A textbook publisher wants to market a new online system for teaching math. As a trial of the system, students in a large number of classrooms are given access to the online activities. In nearly every classroom they see that students who complete most of the on-line exercises later score significantly higher on standardized tests than the students who don't do the exercises. The publisher says "these results show this is a powerful tool for boosting student scores and abilities," but a school administrator says "the results look promising, but this same type of study needs to be replicated on a larger scale nationwide for more credibility."

(a) Do you agree with the publisher? Do you agree with the administrator? Explain.

(b) Was the fact this study was funded by the textbook publisher an example of a confounding factor? Explain.

17. Asthma education. A health insurer would like to reduce the cost of treating asthma through better education of patients. In one city a special course is offered to educate asthmatic patients on how to properly monitor their condition and how to administer medications. The goal is to help reduce mistakes that may require an expensive hospital visit. People who complete the course are followed for two years and insurance records show that on average they require significantly fewer emergency room visits per person compared with asthmatics of the same age and gender in the city who did not take the course.

(a) Does this mean the education program is effective or is there an important confounding factor here?

(b) How could you conduct an experimental study here?

(c) Someone criticizes the study saying that the specific reasons for the emergency room visits need be taken into account to check if they are asthma-related or not. Is this an example of a confounding factor? Explain.

18. Online polls. Yahoo Travel's "best beaches" web page contains an online poll where 74% of the thousands of respondents say that a beach is the most likely place they would visit if taking a leisure trip.[23] Does this mean beaches are one of the most popular destinations for leisure travelers? Or is there an important source of bias in this poll?

[23] `http://travel.yahoo.com/p-promo-3088819`

19. Selling music. A record label would like to see if sending out posters of their artists can generate orders for their records. They choose a large sample of people and randomly send half of them a poster of their most popular artist along with a catalog, while the other half just gets a catalog. They observe that more orders come from the first group than the second group. They conclude the poster is generating the additional sales.

(a) Is this study experimental or observational?

(b) Someone criticizes the study saying "we can't tell if the poster actually motivated people to purchase more music or if those people were going to purchase more music anyway." Is this a potential confounding factor? Is it a reasonable criticism of the study?

20. Music and performance. Researchers would like to see if music can affect people's performance. They randomly divide a very large group of volunteers into two groups: one group gets to listen to loud rap music and the other group listens to soft classical music. The physical strength of each person is tested just after listening to the music. Interestingly, they find that the rap music group is much stronger. They conclude that rap music can make you stronger.

(a) Is this study experimental or observational?

(b) Someone criticizes the study saying "people who tend to listen to rap music tend to be much younger than people who listen to classical music, and this would be a confounding factor." Is this valid criticism of the study?

(c) Someone else criticizes the study saying "you couldn't claim causation unless you also measured everyone's strength before they listened to the music." Is this valid criticism of the study?

More Challenging Exercises

21. Voting felons. Most convicted felons are prohibited by law from voting, but several states are reconsidering these laws. An editorial in *The New York Times* titled "Why Felons Deserve the Right to Vote" argues that the privilege of voting may help ex-offenders become responsible members of society once they are released from prison and may reduce the chance of re-offending.[24] The article refers to data showing that former

[24] "Why Felons Deserve the Right to Vote," *The New York Times*, Monday February 7, 2005, Page A26.

offenders who vote are less likely to return to jail and goes on to say "this lesson has long since been absorbed by democracies abroad, some valuing the franchise so much that they take the ballot boxes right to the prisons."

(a) Does this mean that voting might help reform prisoners or is there an important confounding factor to consider here?

(b) Someone criticizes this study saying social and economic factors play the most important role in whether or not an ex-offender re-offends. Are these examples of confounding factors? Explain?

22. Arbitration effectiveness. An online business-to-business auction firm mediates disputes between client businesses and wants to reduce costs by encouraging clients to settle disputes without a formal hearing. Currently 50% of disputes filed get settled without a formal hearing. They decide to run an experiment to see if they can increase this percentage by providing an optional counseling session before the formal hearing where both sides are encouraged to settle the dispute. They randomly select a large number of disputes filed and offer the clients involved the option to attend a counseling session. Half decide to go to counseling and 90% of their disputes get settled without a hearing. Of the other half that does not go to counseling, the percentage is 40%. Is there evidence here that counseling works?

23. Airline arrivals. An airline's reputation for being on-time is often considered in consumer ratings. Based on one year of data below (also in the file `airlines.xls`) for Alaska Airlines and America West, which of these airlines would you say is doing a better job overall in keeping its flights on time?[25] Justify your answer.

[25]Taken from Arnold Barnett, "How Numbers Can Trick You," *Technology Review*, October, 1994.

	Alaska Airlines		America West Airlines	
Arrival City	# flights arriving on time	# flights arriving late	# flights arriving on time	# flights arriving late
LA	496	63	694	117
Phoenix	220	13	4839	416
San Diego	212	20	383	65
San Francisco	502	103	320	129
Seattle	1841	305	200	62
Totals	3271	504	6436	789

24. Soda and diabetes. A recent newspaper article titled "Study links soda to diabetes in women" describes a high profile study of great concern to executives in the soft drink industry.[26] This study of 91,000 nurses finds "women drinking one or more sugar-sweetened soft drinks a day were twice as likely to develop diabetes as women who drank less than one a month." It goes on to say that pure fruit juice and diet soft drinks were not linked to diabetes, but fruit "punch" (fruit juice with sugar and water added) was linked.

(a) The age of the women was a potential confounding factor, since younger women tend to prefer different drinks than older women. But the researchers properly took this factor into account so it didn't cause bias in the study results. Instead of just comparing all women who drank sugared sodas with all women who didn't drink sugared sodas, what type of comparison(s) would the researchers need to make so that age differences don't bias the study's conclusions?

(b) A soft drink industry spokesman criticizes the study saying that there are some additional important confounding factors that were not completely taken into account. Give an important confounding factor, other than age, that would play a role here.

(c) Explain why the fact that all the subjects in the study were nurses would not be a confounding factor.

(d) Would genetics be a likely confounding factor? Explain why or why not.

25. Cancer-causing drug. The stock price of Teva Pharmaceutical Industries plummeted after results of a study of 892 women came out showing one

[26] "Study links soda to diabetes in women," by Tara Burghart, Associated Press, August 25, 2004

of the company's drugs may be linked to increased risk of breast cancer in women.[27] In defense of the drug, a company spokesman reported that the rate of breast cancer among the 60,000 women who have taken the drug since 1996 was actually lower than the breast cancer rate for women of the same age in the general population. Can you think of a possible confounding factor that would weaken the spokesman's defense? If so, explain why it is a confounder.

26. CFA exam. There are three levels of exams that must be passed in sequence over a number of years to be certified as a financial analyst by the CFA Institute.[28] Part of the prestige of the certification is the difficulty of the exams, which is regularly monitored. The percentage of exam takers who passed the Level I, Level II, and Level III exams in 2004 was 34%, 32%, and 64% respectively. Does this mean the Level III exam was easier than the other two exams or is there an important source of bias here?

27. Best places to work. Each year Fortune magazine ranks companies that are rated highly by employees as a good place to work.[29] The rankings are based in part on surveys sent out randomly to employees, not all of which are returned. Since employees' opinion of their company would affect the chances they return the survey form, is this likely to bias the rankings? Explain briefly.

28. Insiders on fund boards. In 2004 the Securities and Exchange Commission was considering legislation aimed at reducing corruption, which would require that three-quarters of a mutual-fund's board of directors come from outside the company that runs the fund. Critics worried that this requirement may hurt fund performance. A *Business Week* article described a relevant study finding that funds that have board chairmen from outside the company tend to perform worse than other funds that don't.[30] Some experts criticized the study saying that smaller funds are

[27] "Teva Shares Fall on MS Drug and Cancer Paper," Reuters (Chicago), Tue Jan 18, 2005 06:05 PM. See `http://www.reuters.com/newsArticle.jhtml?type=topNews&storyID=7361089`

[28] See `http://www.cfainstitute.org/cfaprogram/cfaprofile/cfa_exam.html`

[29] See `http://www.fortune.com/fortune/bestcompanies`

[30] "Who's Right, The SEC Or Ned Johnson?" by Amy Borrus with Paula Dwyer, *Business Week,* New York: Jun 28, 2004, p. 45. See `http://www.businessweek.com/magazine/content/04_26/b3889048_mz011.htm`

more likely to have chairmen from outside the company and are also more likely to be less profitable (due to overhead expenses) compared with larger funds. How could this confounding factor have been properly taken into account in the study to help better understand the link between having outside chairmen and performance?

29. Perils of benchmarking. A study of one industry finds companies that engage in management practices that are conventionally viewed as risky, such as mandating the use of inter-disciplinary teams, tend to perform better than companies that don't.[31] The study concludes that risky practices are good in that industry.

(a) Do you think this study most likely was observational or experimental?

(b) Describe how survivor bias could have been responsible for the findings.

30. Cell phones and fertility. A recent newspaper article describes a study that finds heavy use of mobile phones may damage men's fertility.[32] It describes how doctors believe damage could be caused by the electromagnetic radiation emitted by handsets or the heat they generate when used close to the body for long periods. The article goes on to say, "the findings suggest millions of men may encounter difficulties in fathering a child due to the widespread use of mobile phones and offers another possible explanation for plummeting fertility levels among British males." Later in the article Dr. Allan Pacey, a senior lecturer in andrology at the University of Sheffield, is quoted as saying "this is a good quality study but I don't think it tackles the issue," and he then mentions several likely confounding factors. Can you think of a few?

31. Immigrants and longevity. A recent *Wall Street Journal* article with the headline "Immigrants Outlive the U.S.-Born Population" describes how black men born in the U.S. live into their 60s on average, black men born in Africa but immigrate to the U.S. live into their 70s on average, but, surprisingly, black men born in Africa who don't immigrate to the

[31] "Selection bias and the perils of benchmarking," Jerker Denrell, *Harvard Business Review,* June 2005.

[32] "Men who use mobile phones face increased risk of infertility," by Jenny Hope, *Daily Mail*, 23rd October 2006, `http://www.dailymail.co.uk/pages/live/articles/news/news.html?in_article_id=412179&in_page_id=1770`. Also see `http://news.bbc.co.uk/2/hi/health/6079782.stm`

U.S. live only into their 50s on average.[33] Since the same type of trend is observed for many other ethnic groups as well, how does immigrating to the U.S. increase your expected lifespan even beyond the average native-born person? Or is there an important confounding factor to consider here? Explain.

32. Quality and IT systems. The health care industry is a large market for information technology (IT), which many believe can both reduce costs and improve quality. A number of hospitals implement new state-of-the-art IT systems and two years later researchers find that the hospitals with the new IT systems had lower cost per patient and higher levels of quality measures than those hospitals without the new IT systems.

(a) Is this an observational or experimental study?

(b) Does this mean the IT systems improve costs and quality or is there confounding factor to consider?

(c) How could you improve the study design without using randomization or forcing hospitals to do things?

33. Cell phones and cancer. A news article describes a study that finds that cell phones do not raise cancer risk.[34] The study looked at all cases of cancer in the entire population of Denmark over a decade and found that people who frequently used cell phones in fact had a much lower incidence of cancer than people who didn't use cell phones. (a) Is this study observational or experimental? (b) Does the lower incidence of cancer mean that using a cell phone may help protect you from getting cancer? Explain.

34. Gender wage gap. The "Fair Pay Act of 2007" is proposed federal legislation to address the troubling fact that, according to the Bureau of Labor Statistics, in 2005 female full-time wage and salary workers earned only 81% of what men did. A newspaper article writes that such a law

[33] *The Wall Street Journal*, May 24, 2004

[34] "Cell phones don't raise cancer risk: study," Wed Dec 6, 2006, by Will Dunham, *Reuters,* `http://today.reuters.co.uk/news/articlenews.aspx?type=healthNews&storyID=2006-12-06T152441Z_01_N05280092_RTRIDST_0_HEALTH-CANCER-CELLPHONES-DC.XML&pageNumber=0&imageid=&cap=&sz=13&WTModLoc=NewsArt-C1-ArticlePage2` On the other hand, a recent news article discusses studies from neighboring Finland and Sweden that find significantly increased cancer risk: see `http://news.independent.co.uk/uk/health_medical/article2472140.ece`

"would bureaucratize most of the labor market" and gives some confounding factors, other than discrimination, that could alternatively explain the salary gap. What factors do you think the article mentions?[35]

35. R&D and profitability. A study of high-tech companies in one industry shows that companies that spend more on Research and Development (R&D) tend to have higher profits: companies devoting 4% of revenues to R&D earned 20% more profit on average than companies devoting 2% of revenues to R&D.

(a) What could explain this type of relationship?

(b) If a company in the industry currently devotes 2% of revenues to R&D, what type of increase in profit might be expected if it increased R&D spending to 4% of revenues?

36. IQ and birth order. A recent study based on the measured IQ scores of 240,000 Norwegian army conscripts finds that "firstborn children are smarter than their siblings—and the reason is not genetics but the way their parents treat them."[36] Researchers attempted to determine whether the IQ difference was caused by genetics or by social interactions within families. Do you have any ideas on how this could be done?

37. Marijuana and psychosis. A study described in a news article finds that "Using marijuana seems to increase the chance of becoming psychotic...even infrequent use could raise the small but real risk of this serious mental illness by 40 percent." To examine the effect of marijuana on mental health, researchers examined 35 studies that tracked tens of thousands of people for periods ranging from one year to 27 years. [37]

(a) Is this study experimental or observational? Explain.

(b) The article also mentions some possible confounding factors. Can you think of one? Justify your answer.

---

[35] "Obama flunks Econ 101," by Cait Murphy, June 5 2007: 7:26 AM EDT, `http://money.cnn.com/2007/06/04/magazines/fortune/muphy_payact.fortune/index.htm?cnn=yes`. See also "The Gender Gap in Wages, circa 2000," June O'Neill, *American Economic Review,* May 2003, Vol.93 No.2, p.309-314.

[36] "Study: Firstborn kids are smarter," by Denise Gellene, *Los Angeles Times*, Friday, June 22, 2007. `http://archives.seattletimes.nwsource.com/cgi-bin/texis.cgi/web/vortex/display?slug=birthorder22&date=20070622`

[37] "Marijuana may increase psychosis risk, analysis says," CNN.com, July 27,2007, `http://www.cnn.com/2007/HEALTH/07/27/marijuana.psychosis.ap/index.html`.

38. Breast implants and suicide. A news article writes "Women who get cosmetic breast implants are nearly three times as likely to commit suicide as other women, U.S. researchers reported on Wednesday."[38] Can you think of a possible confounding factor?

39. Obesity networks. A high profile study of 12,067 people followed for 32 years concludes that obesity can spread from person to person in a social network of friends, much like a virus.[39] One commentator writes that this study only shows that obese people tend to associate with other obese people and not that obesity can be spread through social networks. The researchers, however, had a method for distinguishing the effects of these two different theories. Do you have any ideas on how they did this?

40. Working smokers. A news article writes "a new study shows smokers have poorer-than-average work performance and productivity; they also tend to call in sick more. In a study of more than 14,000 Swedish workers, Petter Lundborg, Ph.D., an economist at the Free University of Amsterdam in the Netherlands, found smokers took an average of almost 11 more sick days than non-smokers."[40] The article goes on to give some potential confounding factors that were adjusted for in the study. Can you think of some?

41. Heart attacks. An observational study concludes that the use of a certain drug lowers the risk of death from heart attacks over a 5-year period. Half the people in the study used the drug and the other half did not. Researchers also asked people in the study about how much exercise they do, whether they smoke, their weight, and their diet. From these data, a single number was assigned to each person ranging from 1 (most healthy behaviors) to 5 (least healthy behaviors).

---

[38] "Breast implants linked with suicide in study," Wed Aug 8, 2007 7:10PM, by Maggie Fox, Reuters, `http://www.reuters.com/article/healthNews/idUSN08369190200708?feedType=RSS&rpc=22&sp=true`.

[39] See "Study Says Obesity Can Be Contagious," by Gina Kolata, July 25, 2007, *The New York Times*, `http://www.nytimes.com/2007/07/25/health/25cnd-fat.html?ex=1187064000&en=6d11916887617c73&ei=5070`, and "The Spread of Obesity in a Large Social Network over 32 Years," Nicholas A. Christakis, and James H. Fowler, *New England Journal of Medicine,* 2007;357:370-9.

[40] "Smokers drag down a workplace, study says,"by Rachel Zupek, CNN.com, August 1, 2007, `http://www.cnn.com/2007/LIVING/worklife/08/14/cb.smokers/index.html`.

(a) The researchers stated they adjusted for differences in healthy behaviors as measured by the scale between those who took the drug and those who didn't. Instead of just comparing people who took the drug and people who didn't take the drug, what type of comparison(s) did the investigators need to make?

(b) A critic claims that the investigators did not adjust for differences in age, and that age would be a serious confounding factor since the risk of heart attacks increases with age. What other information would you need to determine if age was truly a confounding factor?

(c) Later, a randomized experiment was conducted: one group was randomized to receive the drug and another to receive a placebo. After 5 years, people in the treatment group who said they took more than 80% of the pills prescribed had fewer heart attacks than people who took the placebo or who took less than 80% of the pills. An analyst says that since this is an experimental study there are no confounding factors and the conclusion that the drug works – if more than 80% of the pills are taken - is sound. Do you agree with the analyst here? Explain briefly.

42. Television and violence. A newspaper reports the following results of a 17-year study of more than 700 randomly selected families: "Adolescents who watch more than one hour of television a day are more likely to commit aggressive and violent acts as adults...'It's a very important study...it very niftily isolates television as a causal factor' said George Comstock."[41]

(a) Noting that the families were randomly selected, does this mean the study is observational or experimental? Explain.

(b) Someone suspects that gender might be a confounder. Before deciding to adjust for gender, describe how you could check if gender is a confounder.

(c) Someone else suspects that income might be a confounder. Before deciding to adjust for income, describe how you could check if income is a confounder.

43. IT systems and profitability. A study reports a strong positive relationship between a company's profit margin and the extent to which it had a well-implemented state-of-the-art information system. To what extent

[41] *The Los Angeles Times*, March 29, 2002.

is this evidence that a well implemented state-of-the-art information system can improve profit margins? Briefly discuss.

44. Do confident kids do worse in math? A news article titled "Confident Students Do Worse in Math; bad news for U.S." writes "Kids who are turned off by math often say they don't enjoy it, they aren't good at it and they see little point in it. Who knew that could be a formula for success? The nations with the best scores have the least happy, least confident math students, says a study by the Brookings Institution's Brown Center on Education Policy. Countries reporting higher levels of enjoyment and confidence among math students don't do as well in the subject, the study suggests. The results for the United States hover around the middle of the pack, both in terms of enjoyment and in test scores. In essence, happiness is overrated, says study author Tom Loveless."[42] Later in the article it says "Loveless is not suggesting it makes sense to undermine kids' confidence or make math revolting." Based on this study, to what extent does it makes sense to undermine kids' confidence or make math revolting?

45. Autism rates. An article writes "The alarming rise in autism rates in the U.S. and some other developed nations is one of the most anguishing mysteries of modern medicine - and the source of much desperate speculation by parents...Autism may be caused by watching too much television at a tender age...Waldman and colleagues found that reported autism cases within certain counties in California and Pennsylvania rose at rates that closely tracked cable subscriptions, rising fastest in counties with fastest-growing cable."[43] Can you think of any confounding factors?

46. Air-traffic control errors. A news article writes "Air traffic control errors like the one that almost caused two airliners to collide near Chicago this week remain extremely rare and staffing levels are adequate despite controllers' complaints of fatigue and overwork, a federal aviation official said Friday...Evidence points to most of such errors occurring when a

[42] "Confident Students Do Worse in Math; bad news for U.S." POSTED: 7:22 p.m. EDT, October 18, 2006, WASHINGTON (AP), `http://www.ams.org/mathmedia/archive/11-2006-media.html`.

[43] "Does Watching TV Cause Autism?" Friday, Oct. 20, 2006, by Claudia Wallis, *Time.com*, `http://www.time.com/time/health/article/0,8599,1548682,00.html?cnn=yes`.

controller is handling fewer numbers of flights, Cory said, indicating there is no reason to expect an increased likelihood of errors during a hectic air travel period such as Thanksgiving week. She also cited a 2003 FAA study that found 78 percent of errors happen when air traffic is light or average, with only 22 percent occurring when a controller is handling a higher workload of 11 to 15 flights."[44] To what extent do the findings mentioned from the 2003 study shed light on the risks associated with overworked air traffic controllers? Explain briefly.

47. Corporate social performance. A *Harvard Business Review* article under the headline "Do Well by Doing Good? Don't Count on It" writes "Is there, in fact, a link between corporate social performance and corporate financial performance? Not a strong one, according to an analysis of 167 such studies that were conducted over 35 years...we found only a very small correlation between corporate behavior and good financial results."[45] Then the article goes on to propose an alternative reason for the correlation. What do you think the article says?

48. Cell phone radiation. A newspaper article describes a study concluding that the radiation from using a cell phone at bedtime can lead to headaches, confusion and depression because it "causes people to take longer to reach the deeper stages of sleep and to spend less time in them, interfering with the body's ability to repair damage suffered during the day."[46] The treatment group of subjects in the study were exposed to radiation identical to what is received when using a cell phone, while members of the control group were placed in the same setting but received no radiation.

(a) Is this study experimental or observational?

(b) Suppose instead of using cell phone radiation the experimenters simply asked treatment group subjects to talk on a real cell phone near

[44] "FAA reassures travelers after near-miss," by Dave Carpenter, *Associated Press*, Fri, Nov 16, 2007.`http://news.yahoo.com/s/ap/20071117/ap_on_re_us/faa_near_collision_1`

[45] "Do Well by Doing Good? Don't Count on It," by J. Margolis and H.A. Elfenbein, *Harvard Business Review*, Jan 2008, Vol. 86, Issue 1, p19-20.

[46] "Mobile phone radiation wrecks your sleep: Phone makers' own scientists discover that bedtime use can lead to headaches, confusion and depression," by Geoffrey Lean, *The Independent*, January 20, 2008, `http://news.independent.co.uk/sci_tech/article3353768.ece`.

bedtime. Can you think of a likely confounding factor with this design?

(c) The article also mentions that another study "following 1,656 Belgian teenagers for a year, found most of them used their phones after going to bed. It concluded that those who did this once a week were more than three times - and those who used them more often more than five times - as likely to be "very tired." Can you think of a likely confounding factor with this study?

49. Gender and unemployment. A *Business Week* article in May, 2008 discusses data from the Bureau of Labor Statistics and writes "American men are in recession, and American women are not....The share of all men aged 20 and over with jobs has fallen since last November, when private-sector employment peaked, going from 72.9% to 72.2% in April. For women the ratio rose, from 58.1% to 58.3%."[47] What could explain this phenomenon?

50. Charter schools. A news article writes "Charter schools, which are privately run but publicly financed, have been faring well on standardized tests in recent years. But skeptics have discounted their success by accusing them of 'creaming' the best students, saying that the most motivated students and engaged parents are the ones who apply for the spots...Most of the city's 99 charter schools admit students by lottery." The article also describes a well designed study to address this issue. What do you think the design is?[48]

51. Quitting smoking. There is a great deal of interest in determining the effect of a doctor's advice to quit smoking on the behavior of patients who smoke. Researchers study a large representative group of people and compute the percentage of smokers who quit among those who had been advised by their doctor to quit, and then compare this to the percentage among those who had seen a doctor but had not been advised to quit. They see that the first percentage is larger than the second and conclude that a doctor's advice to quit works.

[47] "The Slump: It's a Guy Thing," by Peter Coy, Monday, May 12, 2008, provided by *Business Week,* `http://finance.yahoo.com/career-work/article/105040/The-Slump:-It's-a-Guy-Thing`

[48] Study Shows Better Scores for Charter School Students, By Jennifer Medina, *The New York Times*, September 22, 2009. `http://www.nytimes.com/2009/09/22/education/22charters.html?_r=2&hpw`

(a) The presence of smoking-related illnesses might be a confounding factor because these motivate people to quit regardless of their doctor's advice. What else would have to be true in order for it to actually be a confounding factor? Explain briefly.

(b) The length of time a patient has smoked could be a confounding factor that causes researchers to underestimate the benefit of the advice, because people who have been smoking for a long time find it more difficult to quit. What else would have to be true in order for it to actually be a confounding factor? Explain briefly.

(c) What could be done to reduce the effect of the confounding factor from Part (b)? Explain briefly.

(d) Suppose investigators divide patients into two groups: the first group contains patients of doctors who always advise their smoking patients to quit, and the second group contains patients of doctors who don't always advise their smoking patients to quit. And suppose investigators see that the percentage of smokers who quit was higher in the first group than it was in the second. What else would have to be true for this design to avoid the confounding factors from (a) and (b) above? Explain briefly.

52. Unemployment rates. The graph in Figure 1.4 is from a recent article in *The Wall Street Journal.* The caption says "Jobless rates for workers at each education level have passed peaks reached in recession of the early 1980s. But the combined unemployment rate hasn't eclipsed its prior peak." Though the figure only shows selected education levels, the article mentions that the picture is similar for other educational levels. Since the caption seems to contradict itself, is there an error here or is there another explanation? Explain briefly.[49]

53. College majors. Figure 1.5 from the *Wall Street Journal* shows the results of a survey asking working college graduates if they were satisfied with their career paths. Satisfaction was lowest among psychology majors, where only 26% were "satisfied" or "very satisfied" with their career paths. The article goes on to attribute the low percentage to the fact that "few professions recruit for psychology undergraduate degrees

[49] When Combined Data Reveal the Flaw of Averages, by Cari Tuna, *The Wall Street Journal*, December 2, 2009. `http://online.wsj.com/article/SB125970744553071829.html`

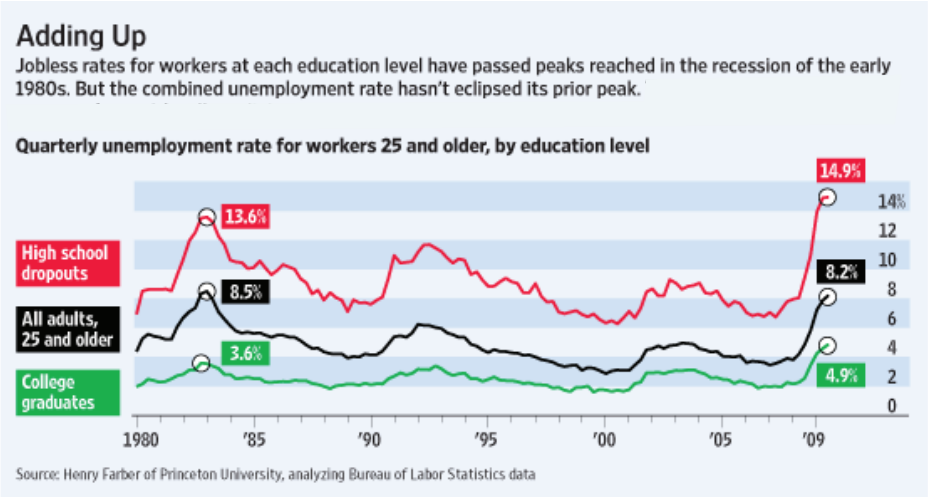

Figure 1.4. Jobless rates for workers at selected education levels from 1980–2009.

specifically" and that "there's not a lot of appetite for their major without a graduate degree." Can you think of an alternate explanation?[50]

54. Driver licensing and crashes. In most US states, teenagers can obtain a driver's license when they turn 16 years old. In response to high accident rates among young drivers, many states have implemented graduated driver licensing (GDL) systems that place restrictions on 16-year-old drivers. For example, 16-year-olds must drive accompanied by an adult for the first 6 months after obtaining their license and then for the next 6-12 months cannot drive at night or with other passengers under age 18. To evaluate its GDL, researchers in one state looked at different age ranges and compared fatal crash rates (the number of accidents that resulted in deaths divided by the number of licensed drivers in the age range) for the 5 years prior to implementation of its GDL and for the 5 years after implementation of the program.[51]

[50] Psych Majors Aren't Happy With Options, by Joe Light, *The Wall Street Journal*, October 11, 2010. `http://online.wsj.com/article/SB10001424052748704011904575538561813341020.html`

[51] See "Graduated Driver Licensing and Fatal Crashes Involving 16- to 19-Year-Old Drivers" by S. Masten, R. Foss and S. Marshall, *Journal of the American Medical Association*, 2011;306(10):1098-110.

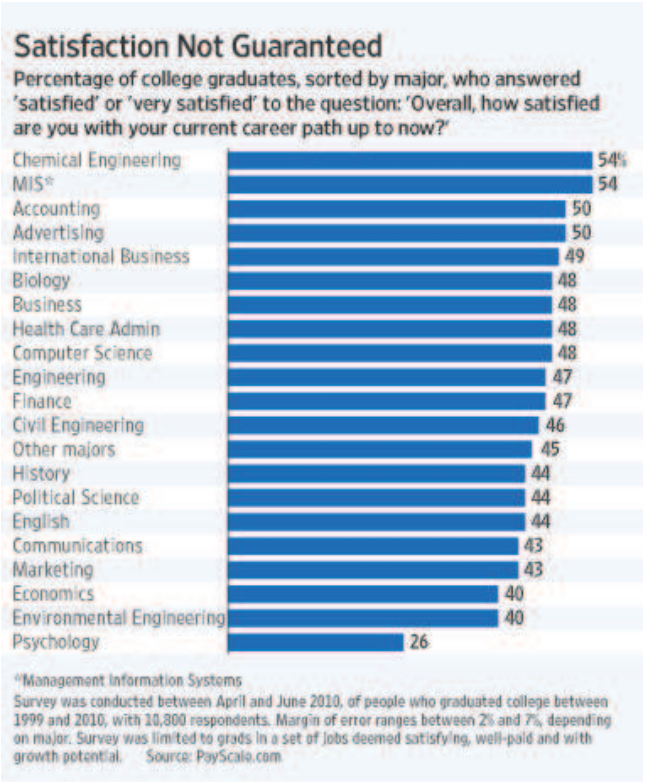

Figure 1.5. The percentage of graduates satisfied with their career for different college majors.

(a) Since the data show there were lower rates of fatal crashes for 16-year-old drivers after implementation of the program compared with before implementation, it looks as though GDL is effective for 16-year-olds. Give an example of a possible confounding factor here and explain in a couple sentences why it is likely to be a confounding factor.

(b) Crash rates for 16-year-olds from another state that had never implemented GDL are also available for comparison for each of the years covered in the study above, but an analyst says that, since rates have always been lower in the other state, the two states are not directly comparable. Could you use the rates from the other state to improve your conclusion about the impact of the GDL sys-

tem? If yes, explain briefly how. If no, explain briefly why not.

(c) Someone criticizes the study saying that the responsibility level of drivers is a likely confounding factor: many 16-year-olds are responsible drivers and would have avoided accidents regardless of whether or not they were accompanied by an adult as part of the GDL requirements. On the other hand, being accompanied by an adult would probably not prevent an accident among the worst 16-year-old drivers. Is this a sensible criticism? Explain briefly.

(d) The results also show that for 18-year-olds there was a higher rate of fatal crashes after implementation of the program compared with before implementation. What do you think might explain this higher fatal accident rates for 18-year-olds? Explain briefly.

55. Noise and heart disease. A study shows that older people exposed to aircraft noise, especially at high levels, may face an increased risk of being hospitalized for heart disease. For seniors living near airports, the study found that every 10-decibel increase in noise from planes was tied to a 3.5 percent higher hospital admission rate for cardiovascular problems. The study also mentions that the researchers adjusted for socioeconomic status. It is also well known that better socioeconomic status tends to be associated with better health. Knowing this, explain why socioeconomic status is likely to be a confounding factor in the aircraft noise study.[52]

56. Does Medicaid hurt or help? A newspaper article discusses a study showing that Medicaid, the health insurance program for the poor, actually hurts its recipients. The evidence given is that Medicaid patients tend to be sicker than the uninsured, and slower to recover from surgery. The article goes on to give an important confounding factor. Can you guess what it is?[53]

57. Obesity, breakfast and television. A news article discusses studies finding that teenagers who eat breakfast regularly and watch less television have less weight gain and lower body mass index when contacted years

[52] Correia AW, Peters JL, Levy JI, Melly S, and Dominici F. Residential exposure to aircraft noise and hospital admissions for cardiovascular diseases: multi-airport retrospective study. *BMJ* 2013;347:f5561

[53] Lousy Medicaid Arguments, by Paul Krugman, *The New York Times*, October 20, 2013

later. The article concludes "encouraging children and teens to eat breakfast and cut back on TV time are important ways to combat obesity." Can you think of some important confounding factors?[54]

58. Anti-anxiety drugs. A news article discusses a study finding that people who regularly took anti-anxiety drugs had more than double the risk of death during the seven-year study period. It mentions that researchers controlled for age, smoking, alcohol use, socioeconomic status, sleep disorders, and anxiety disorders. However, the article says they were not able to control for the severity of the illnesses suffered by the study participants.[55]

(a) Why was it important to control for all those factors? Explain briefly how they could impact the study.

(b) Do you think not controlling for severity of illnesses could impact the study results? If so, explain how.

59. Happiness and sun. A news article titled "Study in Europe eclipses notion home in the sun equals happiness" discusses a study showing that people who migrate from northern Europe to live in warmer climates are marginally less happy than those who stay in northern Europe. The article explains this by saying that, while migrants tended to have higher incomes, "migration itself may be disruptive to other dimensions of people's lives, social ties, sense of belonging, possibly with consequences for their happiness." Can you think of a confounding factor that could be important here?[56]

[54]Eating Breakfast May Beat Teen Obesity, by Jennifer Warner, WebMD Health News, March 3, 2008, `http://www.webmd.com/diet/news/20080303/eating-breakfast-may-beat-teen-obesity`

[55]Anti-Anxiety Drugs Tied to Higher Mortality, by Nicholas Bakalar, *The New York Times,* March 27, 2014 3:57 PM, `http://well.blogs.nytimes.com/2014/03/27/anti-anxiety-drugs-tied-to-higher-mortality/?_php=true&_type=blogs&_php=true&_type=blogs&_r=1`

[56]Study in Europe eclipses notion home in the sun equals happiness, Reuters, April 22, 2014 8:09 PM

Index